

Culpable Aggression: The Basis for Moral Liability to Defensive Killing

Kimberly Kessler Ferzan*

The use of the term, "self-defense," covers a wide array of defensive behaviors, and different actions that repel attacks may be permissible for different reasons. One important justificatory feature of some defensive behaviors is that the aggressor has rendered himself liable to defensive force by his own conduct. That is, when a culpable aggressor points a gun at a defender, and says, "I am going to kill you," the aggressor's behavior forfeits the aggressor's right against the defender's infliction of harm that is intended to repel the aggressor's attack. Because the right is forfeited, numbers do not count (the defender may kill as many culpable aggressors as he needs to), third parties may only aid the defender, and the defender does not owe the aggressor compensation for harms inflicted.

Although liability for culpable aggression seems intuitive, there are a number of questions, including whether culpability is a necessary requirement for liability; what actions the aggressor must perform; whether there must be an actual threat or whether an apparent threat is sufficient; and whether the defender must believe that his use of force is responsive to that threat.

The first part of this paper examines two prominent alternative theories of self-defense—Judith Thomson's rights-based account and Jeff McMahan's moral responsibility account. It argues that both of these theories are problematic as theories of liability and that culpability is a necessary condition of liability. The second part of the paper sketches a culpability account, ultimately arguing:

Liability to Defensive Force:

* Professor of Law, Rutgers University, School of Law—Camden. Associate Graduate Faculty, Rutgers-New Brunswick Philosophy. I want to thank the participants at the University of Missouri Philosophy Department's Kline Workshop on Compensation and Enforcement Rights for the helpful discussion of this paper, and especially, Jeff McMahan for his extensive feedback, and Peter Vallentyne, my commentator, for his extraordinarily thorough analysis. Larry Alexander, Mitch Berman, and John Oberdiek also provided helpful feedback on the draft manuscript.

If:

(1a) *The aggressor forms an intention to purposefully, knowingly, or recklessly kill the defender, the aggressor lacks a justification or excuse, and the defender must kill the aggressor to prevent being killed himself, or*

(1b) *The aggressor purposefully, knowingly, or recklessly engages in conduct that he is aware may lead the defender to believe that (1a) is true, and the aggressor lacks a justification or excuse for so doing;*

Then

(2) *The aggressor is not wronged by force employed by the defender that is intended to repel the perceived attack.*

When one endeavors to give an account of “self-defense,” the initial question is what one means by “self-defense.” As a matter of ordinary language, our use of the term is quite extensive. Certainly, if a Villainous Aggressor tries to kill me, and I kill him, I will justify my action by claiming “self-defense.” The term can also be applied to the killing of an Innocent Aggressor, that is, a person who lacks culpability but whose act will kill me, such as a child who mistakenly believes a real gun to be a toy. And, we can certainly say that one acts in self-defense when one kills an Innocent Threat, an individual who does not even voluntarily act but whose body nevertheless threatens me, such as the fat man who will fall on me and kill me if I do not disintegrate him immediately.¹ Indeed, “self-defense” would likewise be an explanation for why I shot a rabid dog that was about to attack me or maybe even why I disintegrated your television set that was about to smash into me during a tornado. Of course, at some point, my killing or destruction does not require a moral explanation, and at that point, it seems that the use of the term “self-defense” may no longer hold. I think it is at least questionable whether one would argue that one used penicillin to kill bacteria in “self-defense.”² As Bill Murray’s character in *Caddyshack* states (when he learns he is being asked to kill gophers, not golfers), “We can do that; we don’t even need a reason.”³

Given that all instances of self-defense are “defensive” in nature, it is tempting to think that there must be a unifying explanation. I think we ought to resist that temptation. Because we can meaningfully employ the term “self-defense” to justify our killing of a villainous attacker or destroying a neighbor’s television, the explanation would likely have to be so broad as to run roughshod over relevant moral nuances. For these reasons, it is more fruitful to think of self-

¹ These cases were originally imagined by Robert Nozick. ROBERT NOZICK, *ANARCHY, STATE AND UTOPIA* 34–35 (1974).

² For this usage, see Eugene Volokh, *Medical Self-Defense, Prohibited Experimental Therapies, and Payment for Organs*, 120 HARV. L. REV. 1813, 1814 (2007).

³ *CADDYSHACK* (Orion Pictures 1980).

defense as a type of defense, where some tokens may be permissible (or justified), and others will not.⁴

One of the most important normative and conceptual moves in the self-defense literature in the last decade or so was the distinction drawn between permissibility and liability.⁵ Previously, it seemed that one must opt to either take an “aggressor-centered” approach or a “defender-centered” perspective. The assumption was that we needed to pick which approach should govern self-defense. However, this assumption is false. One may believe that both perspectives are morally relevant. A liability approach focuses on what an aggressor must do to be liable to defensive force. A permission-based view focuses on when a defender may harm a person who is, or is perceived to be, an attacker. It is coherent to say that an individual is liable to be killed but the defender lacks the permission to do so, and it is also coherent to claim that an agent may permissibly kill someone who is not liable to be killed.

As one brief example of this severance, consider the well-known trolley problem. It is generally assumed that one may turn a runaway trolley away from five trapped workers to one trapped worker. Many theorists maintain that the one worker’s rights will be infringed, but not violated, because the action is justified. Notably, the worker has certainly done nothing to lose his right to life. Jeff McMahan maintains that in such instances, the worker would have the right to defend himself.⁶ McMahan would say that though it is permissible to turn the trolley, the worker is not liable to be killed. Contrast this with a Villainous Attacker where we would say that the attacker’s rights are neither infringed nor violated. His own actions have forfeited his moral complaint against being killed defensively.

Severing permissibility from liability allows us to see the possibilities in self-defense. I believe it is certainly possible that (1) the moral basis for liability to defensive killing is a culpable attack; (2) a moral permission to kill another may be based on (a) the agent-relative permission to give one’s life greater weight and/or (b) the need for human beings to act on their own perceptions of an attack; and that (3) as a matter of the usage of the term, “self-defense,” it may be appropriate to call actions under either (1) or (2) (or both), instances of “self-defense.”

Although a thorough account of self-defense ought to explore both permissibility and liability and reconcile the two, progress can be made on one front without a focus on the other. Ultimately, liability-based views will have to be

⁴ As suggested by VICTOR TADROS, *CRIMINAL RESPONSIBILITY* 117 (2005); Jeff McMahan, *Self-Defense and the Problem of the Innocent Attacker*, 104 *ETHICS* 252, 256 (1994).

⁵ This suggestion seems to originate in McMahan, *supra* note 4. McMahan has continually distinguished these concepts in his work. It was made quite explicit in Helen Frowe, *A Practical Account of Self-Defence*, 29 *LAW & PHIL.* 245, 245 (2010).

⁶ Jeff McMahan, *The Basis of Moral Liability to Defensive Killing*, 15 *PHIL. ISSUES* 386, 400 (2005) (arguing that defensive force is permitted against both rights violations and infringements). I think this position is misguided, but that need not detain us here.

subsumed under the heading of “permissions,” such that there may be some permissions that are liability-based and some that are not. Liability, that is, will be a ground of permissibility.

In this paper, I will focus on liability. That is, what conditions must obtain for someone to be liable to defensive force? Admittedly, I use the term “liability” somewhat uncomfortably. It is not altogether clear how its usage in the self-defense literature is to be equated with the usage in law or with the usage in Hohfeld’s theory of rights.⁷ But I do believe the term picks out a particularly important notion and is therefore useful as a term of art, however we may choose to cash it out later. (I do some cashing out below.) Ultimately, the question we are asking is what a person must do such that it does not violate or infringe his rights for a defender to kill him.⁸

One further methodological point bears noting. As readers will certainly notice, I am eschewing the justification/excuse distinction in this paper, except insofar as I criticize internal inconsistencies of other theorists. In my view, there is simply no point in battling over the use of “justification” prior to making any progress on the parameters of self-defense.⁹ Rather, what is important here is that we are talking about cases in which the aggressor is not wronged by the defender’s use of force. As I defend below, there are both objective and subjective aspects to liability. Label it as you will.

In the first part of this paper, I will examine two prominent alternative theories of self-defense: Judith Thomson’s rights-based account and Jeff McMahan’s moral responsibility account. I will argue that both of these theories are problematic as theories of liability—that is, forfeiting one’s moral complaint against being killed defensively—and that culpability is a necessary condition of liability. In the second part of this paper, I will sketch a culpability account, including (1) what is the nature of culpability required, (2) how to understand the requirement of aggression, (3) whether there is a requirement that the defensive force be objectively necessary (that is, that the threat be “real”), and (4) whether there is a requirement that the force that repels the attack be intended (or known) to be defensive. In focusing on these features, I am aiming to address various challenges that Jeff McMahan has offered against the culpability account. Specifically, McMahan has criticized my earlier exposition of the culpability account wherein I argued that an individual is liable to defensive force when he plays Russian roulette with an unwilling victim, even if, as it turns out, the bullet was not in the chamber that would have killed the victim, and thus the defense is

⁷ McMahan claims that the notion of liability is a notion from legal theory, and that he borrows from tort and criminal law. *Id.* at 386; JEFF MCMAHAN, *KILLING IN WAR* 44 (2009); *see also* Wesley N. Hohfeld, *Some Fundamental Legal Conceptions as Applied in Judicial Reasoning*, 23 *YALE L.J.* 16 (1913).

⁸ McMahan, *supra* note 6, at 387.

⁹ *See generally* Kimberly Kessler Ferzan, *Justification and Excuse*, in *OXFORD HANDBOOK OF PHILOSOPHY OF CRIMINAL LAW* (John Deigh & David Dolinko eds., 2011).

not “objectively necessary.”¹⁰ In this paper, I will focus only on cases involving deadly force and indivisible harm, thus leaving questions of proportionality for another day.

Ultimately, I will make the case for the following account of “Liability to Defensive Force”:

If:

- (1a) The aggressor forms an intention to purposefully, knowingly, or recklessly kill the defender, the aggressor lacks a justification or excuse, and the defender must kill the aggressor to prevent being killed himself, or
- (1b) The aggressor purposefully, knowingly, or recklessly engages in conduct that he is aware may lead the defender to believe that (1a) is true, and the aggressor lacks a justification or excuse for so doing;

Then:

- (2) The aggressor is not wronged by force employed by the defender that is intended to repel the perceived attack.

I. CULPABILITY AS A NECESSARY CONDITION FOR LIABILITY

One necessary condition for liability to defensive force is *culpability*. The critical question is under what conditions the defender does not violate or infringe the aggressor’s rights by defensively harming him. We might also think that it follows from liability that (1) the attacker cannot fight back and (2) third parties cannot aid the attacker (assuming the defender is not himself liable). It should also follow that the defender would not need to compensate the aggressor for any harm done by the use of defensive force.

There are two prominent views that reject culpability as a necessary condition for liability. The first is the rights-based view proposed by Judith Thomson, and the second is the moral responsibility for an unjust risk view of Jeff McMahan. In this section, I raise objections to both of these views.

A. Examining the Rights-Based View

Judith Thomson argues that one is liable to defensive killing because, by posing a threat to violate the rights of the defender, the aggressor/threat has forfeited his right to life.¹¹ Thomson maintains that even innocent aggressors and

¹⁰ See McMahan, *supra* note 6, at 389–94; Jeff McMahan, *Self-Defense and Culpability*, 24 LAW & PHIL. 751, 751 (2005).

¹¹ Judith Jarvis Thomson, *Self-Defense*, 20 PHIL. & PUB. AFF. 283, 289–91 (1991). Although Thomson does not cast her argument in terms of liability, she focuses on when the aggressor or threat forfeits his right to life, and thus asks the same question with different labels.

threats are rights violators and therefore are liable to be killed. As many have argued, this view is problematic.¹² First, Thomson distinguishes bystanders from threats but fails to do so on any principled grounds. Consider Thomson's *Alcove*:

Alcove. You are in a tunnel and see a runaway trolley headed straight for you and it will kill you if [you] do not escape. You can only escape the trolley by squeezing into a small alcove in the tunnel. Unfortunately for you, there is already someone in the small alcove. You could pull them out of the alcove and onto the tracks, where they will die, so you may fit in the alcove and save yourself.¹³

Thomson rejects that you may pull the bystander out of the alcove.¹⁴ He is morally innocent. But consider:

Fat Man. You have fallen to the bottom of a well. A villain, who has always wanted to kill you, pushes a fat man down the well. If the fat man lands on you, he will live but you will die. You take out your ray gun and disintegrate him.¹⁵

On what basis can we distinguish *Fat Man* from *Alcove*? *Fat Man* is an Innocent Threat, and *Alcove* involves a Bystander. The Innocent Threat and the Bystander are both morally innocent. Moreover, both are causally related to your peril. The Innocent Threat will kill you, and the Bystander prevents your safe use of the alcove.¹⁶ There is simply no principled basis to distinguish the two cases morally.¹⁷

A second question is why it is appropriate to say that Innocent Threats, who are being involuntarily moved, are violating the defender's rights. Though I will

¹² McMahan, *supra* note 4, at 256; Michael Otsuka, *Killing the Innocent in Self-Defense*, 23 PHIL. & PUB. AFF. 74, 74 (1994); Jonathan Quong, *Killing in Self-Defense*, 119 ETHICS 507, 507 (2009).

¹³ As formulated by Quong, *supra* note 12, at 526 (citing Thomson, *supra* note 11, at 291).

¹⁴ Thomson, *supra* note 11, at 289–91 (arguing that the alcove is a “substitution of a bystander” case and that it is impermissible to kill in such instances); see also Otsuka, *supra* note 12, at 76 (arguing for the inviolability of a bystander).

¹⁵ Cf. NOZICK, *supra* note 1, at 34–35.

¹⁶ Larry Alexander, *Self-Defense, Justification, and Excuse*, 22 PHIL. & PUB. AFF. 53, 63 (1993) (noting the reciprocity of risks); Jeff McMahan, *Self-Defense Against Morally Innocent Threats*, in CRIMINAL LAW CONVERSATIONS 385, 389 (Paul H. Robinson et al. eds., 2009).

¹⁷ Quong argues that although there is no basis to distinguish the two, one is permitted to kill the Innocent Threat, but not the alcove occupant, because in the latter case, one uses the Bystander's space and thus violates the means principle. See Quong, *supra* note 12, at 529–30 (arguing it is impermissible to kill another person if you take advantage of, or exploit, his physical space). For a critical review, see Kimberly Kessler Ferzan, *Self-Defense, Permissions, and the Means Principle: A Reply to Quong*, 8 OHIO ST. J. CRIM. L. 503 (2011).

present another way to see this issue below, theorists have argued that it makes little sense to say that he has a duty not to land on you or that he will violate your right not to be killed, though, to be sure, he will kill you.¹⁸

Suzanne Uniacke and Fiona Leverick, both of whom build on Thomson's theory in their monographs devoted to self-defense, have failed to offer any convincing response to this concern.¹⁹ Uniacke, who takes a specification view and argues that one does not have a right to life when one is an unjust threat, does little to explain why the right is so specified.²⁰ Leverick, who adopts a forfeiture view, begins with the claim that we all have a fundamental right to life that we may protect with deadly force but then allows that fundamental right to be forfeited by mere bad luck.²¹ How can something that is sufficiently fundamental that we may kill to protect it, also be lost by mere misfortune such that another person has a right to kill us?²²

Indeed, even if one thinks that it is too hard to ask you not to kill Innocent Threat, the argument is likely to be one of permissibility, not liability.²³ The focus of such arguments is on what we can ask of the defender. One is hard pressed to find a reason why the Innocent Threat is *liable* to be harmed. This becomes even more apparent when we ask what the Innocent Threat may do in response, and whether third parties may intervene on the Innocent Threat's behalf. There is also the question of numbers. Why should a third party intervene to assist you, as opposed to say, 100 babies who are tied together and will land on you if the third party does not disintegrate them?²⁴ This is not the case with 100 Villainous Aggressors. In that case, a third party would not be permitted (barring, say, some special relationship) to side with the numerous attackers over you. Moreover, as Thomson herself noted at one point, it is arguable that the defender owes compensation for killing the Innocent Threat and Innocent Aggressor, a fact that is certainly untrue when one kills a Culpable Aggressor.²⁵

¹⁸ McMahan, *supra* note 16, at 388.

¹⁹ See generally SUZANNE UNIACKE, *PERMISSIBLE KILLING: THE SELF-DEFENCE JUSTIFICATION OF HOMICIDE* (2004); FIONA LEVERICK, *KILLING IN SELF-DEFENCE* (2006).

²⁰ UNIACKE, *supra* note 19, at 194–231.

²¹ Leverick articulates her standard as, “[t]he reason the victim is permitted to kill the aggressor, but the aggressor is not permitted to kill the victim, is that the aggressor, by virtue of her conduct in becoming an unjust immediate threat to the life of the victim that cannot be avoided by any less harmful means, forfeits her right to life.” LEVERICK, *supra* note 19, at 66.

²² Arlette Grabczynska & Kimberly Kessler Ferzan, *Justifying ‘Killing in Self-Defence’*, 99 J. CRIM. L. & CRIMINOLOGY 235, 244 (2009).

²³ Cf. Frowe, *supra* note 5; Quong, *supra* note 12 (both claiming that it is permissible to kill innocent attackers and threats, but not that they are liable to be killed).

²⁴ Alexander, *supra* note 16, at 62 (arguing numbers count with innocent aggressors); see also McMahan, *supra* note 4, at 284.

²⁵ JUDITH JARVIS THOMSON, *RIGHTS, RESTITUTION, AND RISK: ESSAYS IN MORAL THEORY* 41 (William Parent ed., 1986).

B. The Moral Responsibility View

Jeff McMahan offers an alternative account of liability. According to McMahan, the question is whether the actor is morally responsible for an objectively unjust threat of harm.²⁶ If one acts in a way that imposes a foreseeable risk, then one is morally responsible for the threat if that risk materializes.²⁷ Moral responsibility encompasses actions that are subjectively justified—including, say, driving an ambulance and running a red light—because from the objective point of view, the driver should not have acted.²⁸ To be clear, imagine an ambulance driver significantly exceeds the speed limit to get a sick person to the hospital and loses control of the vehicle because of a problem with the power steering. On McMahan's view, if the driver is about to kill a pedestrian, the driver is liable to the pedestrian's defensive force, including, if necessary, deadly force. One is not morally responsible, however, if one does not choose to engage in a known risky activity, even if that activity will cause harm. So, in McMahan's *Cell Phone Operator* case, where an actor's cell phone has been programmed to detonate a bomb that will kill one person but the actor does not know it, McMahan claims that the actor is not liable to be killed because choosing to use a cell phone is not a choice to engage in a risk imposing activity.²⁹ In other words, reasonable risk creation can be a basis for liability. Thus, McMahan finds liability without culpability in *Conscientious Driver*:

A person keeps his car well maintained and always drives cautiously and alertly. On one occasion, however, freak circumstances cause the car to go out of control. It has veered in the direction of a pedestrian whom it will kill unless she blows it up by using one of the explosive devices with which pedestrians in philosophical examples are typically equipped.³⁰

McMahan also sees it as a virtue of his account that *The Resident* is liable to be killed:

²⁶ McMahan, *supra* note 6, at 394 (arguing that under the "Responsibility Account," "the criterion of liability to defensive killing is moral responsibility, through action that lacks objective justification, for a threat of unjust harm to others, where a harm is unjust if it is one to which the victim is not liable and to which she has not consented.").

²⁷ *Id.* at 397 ("I will assume that it is a condition of responsibility for an unjust threat that the action that gave rise to the threat either was of a risk-imposing type or was such that in the circumstances the agent ought to have foreseen that it carried a non-negligible risk of causing a significant unjust harm.").

²⁸ *Id.* at 398.

²⁹ *Id.*

³⁰ *Id.* at 393.

The identical twin of a notorious mass murderer is driving at night in a remote area when his car breaks down. He is nonculpably unaware that his twin brother has within the past few hours escaped from prison in this area and that the residents have been warned of the escape. The murderer's notoriety derives from his invariable modus operandi: he breaks into people's homes and kills them instantly. As the twin approaches a house to request to use the telephone, the resident of the house, reasonably believing himself to be defending his family from the murderer, takes aim to shoot him preemptively.³¹

There are a number of worries about this view. Let us begin with our terms. Consider the relationship between responsibility and liability. Antony Duff's *Answering for Crime* offers an analytical framework for the relationship between responsibility and liability.³² Duff claims that responsibility—which is answerability—is a necessary, but not a sufficient condition of liability.³³ In the criminal law, Duff argues that the actor must perform a culpable act to be criminally responsible, and the actor is only liable if he lacks a justification or excuse. Hence, if Jacob robs a bank, but he acted under duress, then Jacob is responsible for his act, but he is not criminally liable because he is excused. With respect to moral responsibility, Duff argues that responsibility is "strict" but the transition from responsibility to liability still requires not only that the actor lack justification or excuse but also that the actor behave culpably. If Jennifer knocks over my vase, she is responsible for so acting and must answer for it. Her claim, "I didn't see it there," absolves her of liability, but not responsibility.

Now, admittedly, I do not take McMahan to be using "liability" or "responsibility" in this way. He certainly thinks that responsibility is itself sufficient for liability. Duff's framework, however, allows us to question McMahan's assumption. That is, why is strict responsibility sufficient for liability?

Notice McMahan's moral responsibility is strict vis-à-vis his Conscientious Driver and the Resident. McMahan's morally responsible actors are not culpable. They are not even negligent. A conscientious driver, or ambulance driver, who behaves reasonably but nevertheless harms another is not liable in any court of law. Not for even one dollar's worth of damage. It is true that under Duff's framework, the conscientious driver is morally responsible. He is not, however, liable. And indeed, although McMahan suggests that if tort law were governed

³¹ McMahan, *supra* note 16, at 387. Michael Otsuka likewise supports liability here. Otsuka, *supra* note 12, at 91 (actor is liable to defensive force when he sees a realistic hologram of a gun in your hand and aims to kill you defensively because he had rational control over the dangerous activity of using defensive force).

³² R.A. DUFF, *ANSWERING FOR CRIME: RESPONSIBILITY AND LIABILITY IN THE CRIMINAL LAW* 20 (2007).

³³ *Id.* at 20.

solely by moral principles, it would look like his liability view, there is reason to doubt this.³⁴ Although law is certainly not dispositive evidence of moral truths, it does seem to be telling that neither criminal law nor tort law draws the line that McMahan would have us draw.

My objection, however, is not that McMahan's theory is novel. The deeper objection is whether McMahan offers a principled basis for holding some actors who are morally responsible for the risk liable and not others. Why isn't the cell phone operator also morally responsible for the threat he poses, despite his ignorance? What McMahan owes us is a principled account of moral responsibility.³⁵

McMahan's answer to this problem, I take it, is that the conscientious driver and the resident have taken a foreseeable risk. His argument runs as follows: Not every harm we cause results from what we understand to be risk-creating conduct. When it does, however, there is a moral asymmetry between the aggressor who imposes the risk and the defender who does not. Notably, this responsibility is more than causal, as both aggressor and defender are causally responsible for the risk. In other words, a fair reading of McMahan is that moral responsibility is grounded in foreseeable risk imposition.³⁶

Grounding liability in instances of risk imposition that eventuate in threats raises questions about how to treat reciprocal and non-reciprocal riskers. Consider reciprocal riskers first. What happens if we replace the pedestrian with another driver on the road? Among drivers who impose the same risks on each other and make the same choices, there is no principle by which to distinguish them except "moral luck," which appears to be McMahan's answer.³⁷ Moral luck, however, is hardly an answer. Both of them are unlucky—why not say that the defender is unlucky instead of the driver? That is, without some principle that allows the defender to defend, it appears that the defender is the one who is unlucky. And indeed, McMahan rejects that one may kill the Innocent Threat thrown down the well. However, the Innocent Threat is exactly like the driver aggressor. Why does luck play a role in some cases and not others? Although there is admittedly some opportunistic luck involved in questions such as who is located on the street when the car goes out of control, the question McMahan must answer is between individuals who must confront this bad luck: why ought morality to privilege one person over the other? Indeed, consider how a third party should react. Once all

³⁴ McMahan, *supra* note 16, at 395 ("I believe . . . that tort law would correspond more closely to our sense of corrective justice if it were to hold the conscientious driver liable.").

³⁵ Whitley R.P. Kaufman, *Can "Moral Responsibility" Explain Self-Defense?*, in CRIMINAL LAW CONVERSATIONS 400, 400 (Paul H. Robinson et al. eds., 2009) ("[T]he problem is to explain how one can lack culpability and yet have sufficient moral responsibility to be liable to be killed.").

³⁶ Kimberly Kessler Ferzan, *Can't Sue, Can Kill*, in CRIMINAL LAW CONVERSATIONS 398, 398 (Paul H. Robinson et al. eds., 2009); Seth Lazar, *Responsibility, Risk, & Killing in Self-Defense*, 119 ETHICS 699, 711 (2009).

³⁷ Jeff McMahan, *Reply*, in CRIMINAL LAW CONVERSATIONS 404, 405 (Paul H. Robinson et al. eds., 2009) ("among reciprocal risk-imposers, moral luck can determine liability.").

drivers have made the same choices, why should a third party prefer the life of one driver over another? If McMahan would not have a third party intervene to save the man at the bottom of the well, why should a third party intervene to privilege the driver-defender?

Another question is whether assumption of risk affects the driver's moral liability to the pedestrian.³⁸ It is not just that drivers choose to drive cars but pedestrians choose to walk beside them. Both of these choices are reasonable, and both of these choices entail some risk. McMahan does not explain to us why assumption of risk, which does negate liability in tort law, would not negate liability here as well. Tort law's comparative negligence principles complicate the picture further. In one writing, McMahan has suggested that contributory negligence and assumption of risk are relevant.³⁹ This concession substantially complicates the analysis and requires further elaboration than McMahan has given.

Moreover, even in cases that initially appear to involve nonreciprocal risk imposition, there are risks being imposed in both directions. A pedestrian who has a gun and shoots the Conscientious Driver in self-defense was imposing a risk by carrying around a gun.⁴⁰ Not only is there the risk the gun will go off accidentally, but also the pedestrian's carrying a gun and being present makes it more risky for the driver to drive.⁴¹ That is, in almost any case in which self-defense will be successful (that is, that does not rely on the defender just finding a weapon when he needs it), the defender will be imposing a risk on the driver. Indeed, imagine a pedestrian plants mines along the side of the road so that they will explode if a driver veers out of control, and the next day, the conscientious driver drives and the pedestrian walks.⁴² If the conscientious driver's car suddenly malfunctions, why is it that he is the one liable to defensive force and not the pedestrian? The pedestrian has made the driver's driving riskier, and he has imposed a risk on the driver, a risk that now will result in a harm. So, the pedestrian appears equally liable to defensive force.

Even assuming the risks are not reciprocal and that we can ignore assumption of risk, there are reasons to reject McMahan's liability conditions. Consider the requirement that the harm must be foreseeable. Foreseeability will always be a matter of (1) the selection of the description of the harm and (2) the selection of the information available to the assessor. (All harms that occur are foreseen by the

³⁸ Lazar, *supra* note 36, at 718; see also JUDITH JARVIS THOMSON, *THE REALM OF RIGHTS* 180 (1990) (noting that with respect to the trolley problem, it matters why the five came to be in the trolley's path in determining whether it is permissible to turn the trolley).

³⁹ McMahan, *supra* note 16, at 405.

⁴⁰ Lazar, *supra* note 36, at 719; see also Kimberly Kessler Ferzan, *Justifying Self-Defense*, 24 *LAW & PHIL.* 711, 736 (2005) (noting the Coasean objection to McMahan's conditions of liability); see generally Ronald H. Coase, *The Problem of Social Cost*, 3 *J.L. & ECON.* 1, 2 (1960) (noting causation is reciprocal).

⁴¹ Lazar, *supra* note 36, at 719. I also thank Larry Alexander for the latter point.

⁴² I thank Larry Alexander for this example.

omniscient, even the cell phone.) It is true that driving seems risky (people die in car accidents) but it isn't true that we would say that safe driving with a well-maintained vehicle is risky vis-à-vis a mechanical malfunction. That is, the question, "what if my well maintained brakes fail?" is not the sort of risk the Conscientious Driver foresees when he chooses to drive his car.

Moreover, does McMahan require that the harm be foreseen by the actor? Would it matter if Fearless Fred never thinks that any harm can come from driving his car? In contrast, McMahan argues that the Innocent Threat's threat is not one he "could conceivably foresee."⁴³ However, what if Nervous Nellie reads the literature on self-defense and becomes convinced that we are in the midst of an epidemic whereby evil villains throw people down wells? Will her mere appearance in public now be a risk-imposing activity because she believes it is? Are we to determine whether the risk is foreseeable from the standpoint of the agent, a fully objective standpoint, or some place in between?⁴⁴ As to the third possibility, if the governing perspective is not the actor's then how does McMahan justify importing some artificial epistemically limited perspective within his objective test? And importantly, how can morality speak through a construct that mirrors neither the actor's perspective nor the omniscient's?⁴⁵

⁴³ MCMAHAN, *supra* note 7, at 167–68.

⁴⁴ In private conversation, McMahan suggested that he would apply Parfit's distinction between types of justifications as fact-relative (what criminal law theorists generally see as objectively justified), evidence-relative (what one should do relative to the information base one has), and belief-relative (what is the fully subjective agent's point of view). McMahan would therefore say that the Conscientious Driver is unjustified in the fact-relative sense, while justified in the other two senses. But conceptual distinctions cannot alone solve normative dilemmas. That is, Conscientious Driver and Cell Phone Operator are both fact-relative unjustified and evidence-relative justified. It may be true that the Conscientious Driver *knew* that there could be a risk of harming someone while driving and the Cell Phone Operator did not have the evidence available to make that assessment, but I still do not see how and why anything turns on that.

Adding to the dilemma is that in a recent draft manuscript, McMahan appears to have at least partially abandoned the requirement that the person actually present a threat, and may have finally come around to the view I have endorsed since 2005 (and discuss *infra*) that apparent threats are also liable to defensive killing. Jeff McMahan, *Who Is Morally Liable to be Killed in War*, 71 ANALYSIS REVIEWS 544 (2011). McMahan does not say exactly that, but does note that the defender is not liable to counterattack for killing an apparent threatener even if the threatener does not actually pose a threat. Interpreted modestly, McMahan endorses a provocateur forfeiture rule that someone who (culpably) causes the conditions of his defense cannot thereby engage in it. Thus, if I lead you to believe I am attacking you when I am not, then when you defend yourself, I cannot defend against you, as I caused the affray to begin with. But, read broadly, McMahan could potentially be endorsing the view that one need only appear to be a threat and need not be a threat in the fact-relative sense. Either way, this concession places significant emphasis on evidence-based and belief-based senses of justification and moves us further away from the fact-relative sense (because an apparent threatener will not harm you viewed from that perspective). McMahan has yet to explain which view matters when and why.

⁴⁵ Heidi M. Hurd, *Nonreciprocal Risk Imposition, Unjust Enrichment, and the Foundations of Tort Law: A Critical Celebration of George Fletcher's Theory of Tort Law*, 78 NOTRE DAME L. REV. 711, 722–23 (2003).

By employing foreseeability and risk criteria, McMahan oddly tries to straddle the objective/subjective divide. Doctrines like foreseeability and probability require that we select some epistemic stance, whereas objective justification does not. McMahan sometimes conjoins the two. In *Killing in War*, McMahan states, “[a]lthough the threat posed by ordinary drivers is normally so low that no defensive action against them could be proportionate, they can become liable to defensive action when the threat they pose passes a certain threshold of probability.”⁴⁶ Notice, the reason McMahan thinks the Conscientious Driver and the ambulance driver are not justified is because, from an omniscient point of view, their actions will cause harm. Neither should drive. This privileges a purely objective standpoint. On the other hand, McMahan’s reliance on probabilities suggests that we are no longer looking at this question from the perspective of objective justification. If we are taking this latter approach, which is a weighing of risks and reasons for action, then could a driver who will cause harm not be liable because from an epistemic perspective, the probability is not high enough? Conversely, why is the driver who will not cause harm not liable when both he and the pedestrian predict he will? I do not see how McMahan can allow probabilities to enter into his liability conditions if he believes that justifications are objectively determined.

It is also strange that McMahan allows some epistemic errors but not others to render one liable. According to McMahan, the reason why the Conscientious Driver may be killed is that he engages in an activity which (1) he knows will impose a risk even though (2) he reasonably believes the risk is justified. The cell phone operator’s epistemic error occurs earlier—he does not know the phone will detonate a bomb. But, why should it matter? From an objective perspective, both individuals will cause harm. From a subjective perspective, both actors behave reasonably. Why if one is subjectively justified in imposing foreseeable risks does one become liable to be killed but not another person who is ignorant of the fact that his conduct also imposes a subjectively justified risk? In addition, such a choice might make more sense if we lived in a world in which we could choose not to impose risks, but we don’t. If Alice is poor and must drive a significant distance to get to work, she has every reason to drive. She just doesn’t know that one day she will kill a pedestrian when her car goes out of control. But it seems hard to say that Alice’s choice renders her liable to be killed when she does everything that we can fairly ask her to do. It is not that Alice cannot control what happens, but that it

⁴⁶ MCMAHAN, *supra* note 7, at 43–44. To further beat the horse and truly hope it is dead, I find this quotation extraordinarily perplexing. Which sense of justification does McMahan mean to employ here? It must be evidence or belief-based because from the fact perspective, the probability is either 0 or 1. And, if we are speaking about evidence-based justification, then, from whose perspective are we assessing the probability? Have we now shifted from the aggressor to the defender? They aren’t likely to have exactly the same evidence. Further confusion comes when McMahan continues the argument and claims that drivers are not likely to be justified, but just permitted, to drive, given the reasons they have. *Id.* at 44. But that sounds like fact-based justification and certainly not an evidence-relative justification based on the defender’s perspective.

is unfair to ask her not to engage in the conduct. Similarly, a Fat Man could certainly keep from being thrown down a well by locking himself inside his home, but we cannot fairly ask that of him either.

Indeed, let me gesture at an argument against the conventional wisdom that the Innocent Threat is a nonresponsible threat. Assume that Albert is told by Victor, "I want to kill Betty, and, Albert, if I ever get the chance, I am going to shove you on top of her down a well." I take it that we would not say that Albert imposes a risk on Betty simply by walking down the street. Now, the conventional wisdom is that Albert lacks control over the threat he poses to Betty, but that is not quite right. If you were told, "If you ever kick me, I will kill your mother, father, and every other person near and dear to you," you would find a way to never, ever kick me. Not purposefully. Not by accident. Not by reflex. You would engage in myriad behaviors, many excessive, to prevent yourself from ever kicking me. You might, for example, simply cut off your legs. The critical point would then be not that you could not control kicking me, but that, as a matter of claims that I have against you, it is not fair for me to ask this of you.⁴⁷ So, it is not that Albert cannot control being hurled on Betty, but that it is unfair to ask Albert to so control.

Once we move from nonresponsibility to unfairness, however, we see that many of McMahan's responsible threats are being asked to shoulder unfair responsibilities. The mere foresight that harm may occur from an activity should not be sufficient to make you liable when you do all that we may fairly ask of you.⁴⁸ The conscientious driver does just that. He behaves conscientiously.⁴⁹

McMahan ultimately concedes that whatever asymmetry there may be between his Responsible Threats and the victim is slight.⁵⁰ However, particularly in cases where harms are not divisible and the sort of defensive force that must be employed is lethal, a mere slight asymmetry between the defender and the aggressor should not be sufficient. The aggressor must perform an action that we can fairly say forfeits his moral complaint against defensive force being applied

⁴⁷ DUFF, *supra* note 32, at 60; see also Otsuka, *supra* note 12, at 81 (arguing that insofar as we have intuitions that one may kill an Innocent Threat it is because we assume Innocent Threat could have taken precautions against becoming a threat).

⁴⁸ I should note that I am highly dubious of basing liability to defensive force or criminal liability on negligence. See, e.g., LARRY ALEXANDER & KIMBERLY KESSLER FERZAN, *CRIME AND CULPABILITY: A THEORY OF CRIMINAL LAW* 69–85 (2009). Thus, my argument that McMahan's theory is woefully overinclusive should not be read as an endorsement of negligence *per se*.

⁴⁹ For this reason, I would also reject Michael Otsuka's claim that cases on par with *The Resident* are instances of liability because the mistaken self-defender "takes a gamble by placing this moral immunity on the line when she engages in such avoidable risky activity." Otsuka, *supra* note 12, at 91. I am not sure how mistaken self-defenders who honestly and reasonably believe they are being attacked are taking gambles and engaging in avoidably risky behavior. After all, they are likewise taking a gamble if they do not defend themselves, and they are gambling with their lives. I do not see how or why we should say that such actors forfeit their rights to life when they do all that morality can fairly ask of them.

⁵⁰ MCMAHAN, *supra* note 7, at 178.

against him. It is still possible that both sides may *permissibly* act in their own defense, though neither is liable.⁵¹

II. THE CONDITIONS OF CULPABILITY-BASED LIABILITY

What the previous views share with my own is that culpability appears to be sufficient, even if not necessary, criterion for liability. A culpable attacker meets both Thomson's and McMahan's criteria.⁵² However, even if it seems intuitively obvious that culpable attackers are liable to be killed defensively,⁵³ there are a number of nuances to be worked out. In this section, I will sketch a culpability-based view. In so doing, I will respond to a number of concerns raised by Jeff McMahan.⁵⁴

Where the prior views substantially depart from my own is with the requirement that there be an actual threat. To Thomson, it is the fact that the threat will kill you that allows you to kill him. To McMahan, only once the act is objectively (fact-relative) unjustified, do we look to see if the actor is liable to defensive force. I believe this is misguided because it ignores the preemptive and preventive nature of self-defense. Namely, when an aggressor is liable to self-defense, he is not wronged by the defender's acting *on the prediction* that she will be harmed.

As mentioned above, I will make the case for the following conditions for liability:

If:

- (1a) The aggressor forms an intention to purposefully, knowingly, or recklessly kill the defender, the aggressor lacks a justification or excuse, and the defender must kill the aggressor to prevent being killed himself, or
- (1b) The aggressor purposefully, knowingly, or recklessly engages in conduct that he is aware may lead the defender to believe that (1a) is true, and the aggressor lacks a justification or excuse for so doing;

Then:

- (2) The aggressor is not wronged by force employed by the defender that is intended to repel the perceived attack.

⁵¹ Cf. *id.* at 180.

⁵² See *id.* at 159.

⁵³ Many theorists do not think we even need an account vis-à-vis the culpable. See Quong, *supra* note 12, at 523 (describing the permissibility of killing a Villainous Aggressor as "a relatively unproblematic case where someone forfeits their right not to be killed").

⁵⁴ McMahan, *supra* note 6, at 389–94 (criticizing the culpability account); Jeff McMahan, *Self Defense and Culpability*, 24 LAW & PHIL. 751 (2005).

A. Culpability Generally

The first question is what I mean by culpability. Typically, in the case of self-defense, the defender will be acting against someone who intends to kill him. Requiring specific intent, however, is too narrow. Rather, it seems that when an aggressor acts purposefully, knowingly, or recklessly (and without justification or excuse), any of these states should suffice for altering the culpable aggressor's status vis-à-vis the defender such that the defender does not wrong the culpable aggressor by killing him.

That is, the criminal law's model of culpability best captures the sort of culpability inherent in culpable aggression. Intending to kill the victim as an end in itself, or as a means to this end, is clearly culpable. One needs little argument to say that attacking with such an intention should be a sufficient mental state for liability. Knowledge should also suffice. Clearly, if one is a passenger on a plane, the fact that the "attacker" plans to blow up the plane for the insurance money on the plane, as opposed to killing the passenger, makes little difference in terms of the attacker's liability to defensive force. Notably, in instances of both purpose and knowledge an actor is not culpable if his conduct is either justified or excused.

Recklessness should also be a sufficient mental state. Recklessness requires that an attacker consciously disregard a substantial and unjustifiable risk to the defender. Recklessness includes a person who plays involuntary Russian roulette, a person who drag races along a busy street, and other persons who impose risks on the defender for insufficient reasons.

Elsewhere Larry Alexander and I have argued that there is one unified theory of culpability, with recklessness roughly expressing this form.⁵⁵ Culpability consists in creating a risk for insufficient reasons. And, although purpose (intention) looks different from this, purpose is actually acting for a presumptively insufficient reason. This presumption, however, can be rebutted when the actor is justified (for instance, acting in self-defense). Knowledge is also about a risk/reason comparison where the risk appears to reach certainty. However, one is ultimately not culpable if one knowingly harms another for a justifiable reason. Even if one rejects our view that these mental states roughly collapse into recklessness, one may still agree that purpose, knowledge, and recklessness are culpable mental states that are sufficient to satisfy the mental element of culpable aggression. An actor who intends to engage in an action that demonstrates insufficient concern for the interests of other people can hardly complain if that action is stopped.

Importantly, our view likewise sees culpability broadly, as including an evaluation of whether the aggressor is excused or justified, in determining whether the aggressor is culpable. That is, an intention to kill that is morally justified is not a culpable attack. And, an intention to kill that is the product of duress, insanity, or mistake, is also not a culpable attack.

⁵⁵ ALEXANDER & FERZAN, *supra* note 48, at 69–85.

Thus far I have said nothing about negligence. I reject negligence as sufficient for criminal blameworthiness, and I would likewise reject it here.⁵⁶ Individuals may fall below the “reasonable person test” because of moments of stupidity, clumsiness, or bad character.⁵⁷ I do not believe that we should be able to forfeit our right to repel a deadly attack, and our moral complaint against such an attack, based upon momentary failures. If *Unconscientious Driver* hears that her brakes are squeaking and thinks “oh, that is nothing” when it is a problem with her brakes, it may be the case that she has some sort of character failure. However, this failure strikes me as insufficient for forfeiting one’s right. A full defense of this view, however, requires a paper in its own right and nothing I say here turns on it. If you believe that negligence is culpable, then you can simply include it in your culpability account.

In all cases of culpability, the aggressor has clearly altered his moral standing vis-à-vis the defender. The aggressor intends to violate the defender’s rights. Because this is something the aggressor is not permitted to do, his culpability renders him liable to defensive force.

At this point, the potential objection is that the view is under inclusive.⁵⁸ If culpability is a necessary element of liability to defensive force, then the Conscientious Driver, the Resident, and the Innocent Threat are not liable. I believe this to be true. Liability to defensive killing should require a significant showing. Most importantly, liability is not about when the defender needs to act, but when the aggressor has forfeited his rights such that it is not wrong to use defensive force against him. We may believe that the pedestrian should be permitted to kill the Conscientious Driver because, say, he has an agent-relative permission to prefer his life to the driver’s, but this is not to say that the Conscientious Driver has engaged in the sort of behavior whereby he should be deemed to have forfeited his defensive rights.

At this point, the formulation first appears to be that:

When *A* will purposefully, knowingly, or recklessly kill *D*, and *A* lacks justification or excuse, *A* is liable to defensive force by *D*.

This, however, is just a first pass. We must attend to the question of what an aggressor must *do* in order to be liable to defensive force.

⁵⁶ See *supra* note 48.

⁵⁷ See generally Michael S. Moore & Heidi M. Hurd, *Punishing the Awkward, the Stupid, the Weak, and the Selfish: The Culpability of Negligence*, 5 CRIM. L. & PHIL. 147 (2011).

⁵⁸ On this point see McMahan, *supra* note 6, at 393 (arguing culpability is not necessary for liability).

B. Aggression's Act Requirement

Free-floating culpability, however, will not cut it. When we think about a culpable aggressor, we think about the aggressor *doing something* that renders him liable to defensive force. The question of how to specify what aggression consists in turns out to be a complicated task. Indeed, in this section, I will not aim to resolve fully the "test" for aggression. I wish only to make the case that we do need such a concept, and to gesture at where we might begin to find the answers.

We should not be surprised that this is a difficult task. With respect to attempts, the criminal law has struggled quite significantly with the act requirement. Where does preparation end and attempting begin? Is an act only meant to corroborate the evil intention? Should the act requirement be placed close to completion so as to allow for renunciation? Although some theorists maintain that the intention itself is sufficient for the punishment (and the act is merely evidentiary), other theorists argue that we should wait for the last act so as to ensure that we are punishing the actor for what he has done and not what he will do.

Any account of self-defense requires an account of the action that the aggressor must take to trigger the defender's defensive right. Consider the following.

Fortune Teller. When Vera is ten years old, Fortuna tells her that at some point in her life, Ivan, who will be born in five days, will pose a threat to her life and she will need to kill him to prevent him from killing her. Vera moves to a different town. Twenty-five years later, when traveling on vacation, she sees Ivan at a store. Two days later, Vera trips and falls into a well. Vera's twin sister, Vena, who has always hated Vera, sees Ivan walk by the well and shoves him down the hole. Ever since her fortune was told, Vera has brought a ray gun with her so she can disintegrate Ivan and does so.

If Vera has a right to kill Ivan because he will otherwise violate her right not to be killed, that fact is true, not just once Vena pushes Ivan down the well, but two days prior, and even twenty-five years prior. It appears that Thomson's view would justify Vera killing Ivan at birth (or perhaps *in utero*).⁵⁹ Now consider the actus reus issue with respect to McMahan's theory:

Total Recall. Dena is driving a Loyota, a car brand that has recently been in the news for a failure of power steering. Victor is an

⁵⁹ Russell Christopher convincingly argues that without a subjective element that distinguishes attacks from responses, the defender is liable to defensive force before the aggressor is. See Russell Christopher, *Self-Defense and Objectivity: A Reply to Judith Jarvis Thomson*, 1 BUFF. CRIM. L. REV. 537, 539 (1998).

investigative reporter who is about to publish an article on how all red Loyotas have this problem and the power steering fails two minutes after the driver pumps his brakes twice quickly. As he is walking through a tunnel, he sees Dena's red Loyota and sees Dena stop quickly for a red light by pumping the brakes twice. In this tunnel, Victor is certain (and correct) that once Dena loses control of the car, she will come right toward the pedestrian walkway, swerve toward him, and, without intervention, kill him.

McMahan's view is that once one chooses to engage in risky behavior, one is liable to defensive force if the activity turns out to be one that will impose harm. Now, McMahan may be able to prevent a total pushback along the time line, as he may be able to argue that until Dena chooses to drive the car, she is not morally responsible for the threat (as most criminal law theorists and moral philosophers create room for compatibilism), but her responsibility for the threat certainly obtains at the moment she gets in the car, as opposed to at the later moment for the brake failure. So, it would seem that Dena is morally liable to defensive force by Victor even prior to the power steering failure.

Hence, any theory of self-defense will have to have a theory of aggression that specifies how far the attacker must come. In asking this question, we must bear in mind that we are dealing with *liability* and not just *permissibility*. We are asking the conditions under which an aggressor gives up a right not to have defensive deadly force used against him. This idea may (but does not necessarily) come apart from how much risk it is fair to ask the defender to bear.

Now, one possibility is that once the aggressor has formed a culpable intention, then the defender is permitted to act when she needs to. But "need" is a dangerous category. A defender may need to act prior to the formation of a culpable intention by the aggressor or the acquisition of any weapons necessary to launch an attack. If the right to self-defense is simply a right to act as early as is necessary to defend oneself effectively, then the need to defend may very well arise far before the initiation of any aggression.⁶⁰

Indeed, consider how McMahan tortures his own view of liability to get to need-based cases. In *Killing in War*, McMahan imagines soldiers who join their country's army for good reasons. Their government is plotting an unjust war against a neighboring country, but the soldiers are oblivious to this fact. May the neighboring country preemptively attack the soldiers? McMahan thinks so. "They earlier made a voluntary choice that in effect committed them in a public way to obedience, and those to whom they owe obedience will, unless prevented, order them to fight in an unjust war in which it is reasonable to expect they will

⁶⁰ See Kimberly Kessler Ferzan, *Defending Imminence: From Battèred Women to Iraq*, 46 ARIZ. L. REV. 213, 250–51 (2004) (noting the need operates independently of the aggressor's intention and current capabilities).

participate.”⁶¹ I find this result troubling. The soldiers have not chosen to engage in any risk imposing activity, nor have they been given the opportunity to refuse to engage in an unjust war. I do not see why the soldiers have, at this point, forfeited their moral complaint.

One might think that the culpability view solves this problem because it requires *need* and culpability. Let us detangle three different ideas. First, agents may have agent-relative permissions to prefer their lives to others’ when they *need* to. This idea does not require culpability. Second, in cases where one person needs to die, an individual’s culpability in creating the situation is a significant consideration.⁶² Indeed, it may be that as a matter of selection, life’s worth and overall culpability ought to matter.⁶³ Third, if someone intends to kill you and you need to act at the moment of intention formation to stop him, you may act then. This third proposition is deceptively simple. As Larry Alexander and I have often argued, there are a number of problems with holding people criminally liable for early attempts—particularly an act like intention formation.⁶⁴ It is difficult to distinguish intentions from fantasies; intentions are often internally and externally conditional; and intentions are revocable. Alex may intend “to kill Betty” but change his mind when he sees her smile (an externally conditional intention). Or Alex may intend “to kill Betty unless she smiles” (an internally conditional intention). Or Alex may intend to kill Betty, and then just simply think the better of it. Larry Alexander and I have argued that we cannot punish Alex based on the prediction that he will act because he very well may not.

Now, one way out of this dilemma for self-defense is to argue that it must be the case that Alex would have killed Betty for Alex to be liable. If Betty makes a mistaken prediction, she may not be punished (and legal theorists can debate whether a reasonable mistake makes her justified or excused), but Alex is not liable because he was not really going to do it. The thought is that “need” solves the dilemma. That is, we ignore all problems with “what if the aggressor would have changed her mind” by presupposing, with “need,” that she won’t. If we are entitled to assume necessity, then it appears that there should not be a requirement of action. That is, between an aggressor and a defender, if the defender must kill the aggressor, he is entitled to do so when the aggressor has formed what is, at that time, a culpable intention, and will remain a culpable intention in its execution. Between the innocent defender and the culpable (future) aggressor, the aggressor’s

⁶¹ MCMAHAN, *supra* note 7, at 183.

⁶² Phillip Montague, *Self-Defense and Choosing Between Lives*, 40 PHIL. STUD. 207, 211 (1981) (arguing that in forced choice situations, “special moral significance” does attach to the situation’s being caused by someone’s culpable behavior); David Wasserman, *Justifying Self-Defense*, 16 PHIL. & PUB. AFF. 356, 370 (1987) (rejecting Montague’s theory and arguing “that greater fault is not a sufficient condition for choosing one life over another”).

⁶³ ALEXANDER & FERZAN, *supra* note 48, at 129–31 (discussing the range of possibilities).

⁶⁴ See *id.* at 197–225; Larry Alexander & Kimberly Kessler Ferzan, *Danger: The Ethics of Preemptive Action* 9 OHIO ST. J. CRIM. L. 637 (2012); Larry Alexander and Kimberly D. Kessler, *Mens Rea and Inchoate Crimes*, 87 J. CRIM. L. & CRIMINOLOGY 1138 (1997).

own choice is sufficient to alter his moral standing and allow defensive force to be used against him.

We might revisit our liability formulation thus:

When *A* will purposefully, knowingly, or recklessly kill *D*, and *A* lacks justification or excuse, *A* is liable to defensive force once *A* forms a culpable intention and *D* must kill *A* to prevent being killed himself.⁶⁵

C. Objective Necessity

Still, there is something altogether odd about premising the liability conditions of the aggressor on whether it will in fact be necessary to kill him—a fact that the defender cannot know, but can only predict. In the punishment context, this raises the question of pre-punishment.⁶⁶ That is, we might worry about someone deserving something before his criminal act *actually* happens. In the context of self-defense, the concern is different. The concern is that a moral principle that relies on necessity ignores the very sort of practice that self-defense is. Self-defense can never be about whether the aggressor would have succeeded because self-defense is a moral permission to preempt an anticipated harm.

That is, I think that we can no more make sense of the question of whether we can pre-punish someone for a criminal act he has not done, then whether we can justify preempting a harm that definitely will occur. Both questions are confused. Punishment is responsive to an offense committed. Prevention is responsive to a threat perceived, and thus, for self-defense, the question should be what the offender must do that creates such an appearance such that he is liable to defensive force.

In other words, objective necessity is not a requirement of liability. Consider:

⁶⁵ There may be reasons, even when a killing is necessary, to delay the killing until the last minute at which the defender's action will still be successful. Gideon Yaffe argues that the imminence requirement is necessary because it ensures that the defensive action only occurs in a possible world that is very close to the one in which the aggressor *deserves* punishment. Gideon Yaffe, *Prevention and Imminence, Pre-Punishment and Actuality* 48 SAN DIEGO L. REV. 1205, 1220–21 (2011). Although Yaffe is mistaken to claim that the fit between forfeiture and prevention is the same fit as between wrongdoing and desert, his analysis is insightful in the way that it claims that an aggressor cannot see himself as committing the culpable act until very close to its commission. Whether this would require that the defender wait until the threat is imminent is beyond what we need to address here. I will argue presently that liability can be premised on appearances, and not necessity.

⁶⁶ See *id.*; see also Christopher New, *Time and Punishment*, 52 ANALYSIS 35 (1992) (introducing the puzzle); Christopher New, *Punishing Times: Reply to Smilansky*, 55 ANALYSIS 60 (1995) (arguing that waiting is just an “empty gesture” if we know the defendant will commit the act); Saul Smilansky, *The Time to Punish*, 54 ANALYSIS 50 (1994) (arguing that pre-punishment does not show respect for persons); Daniel Statman, *The Time to Punish and the Problem of Moral Luck*, 14 J. APPLIED PHIL. 129 (1997) (arguing that New should be willing to punish even prior to the formation of an intention to commit a crime).

Involuntary Roulette. Greta puts one bullet in a six-chamber revolver, spins the cylinder, and points the gun at Harry. Assume that Greta will only take one shot and Harry knows this. Harry shoots Greta. The bullet was not in the fateful chamber.

Unloaded Gun. Ike runs into a liquor store with a gun. After the clerk gives him the money, Ike points the gun at the clerk and says, "Now, I am going to blow your head off." The clerk shoots Ike. As it turns out the gun was unloaded—as Ike knew—he just wanted to toy with the clerk.

Recently, theorists have begun to suggest, with mere comments in passing, that there could be liability for bluffs.⁶⁷ I, too, believe that we ought to regard these as instances of liability,⁶⁸ but we must first confront a powerful objection to the view. The challenge, as Jeff McMahan succinctly puts it, is to explain "how your lack of morally relevant knowledge can establish a justification where otherwise there would be none."⁶⁹ Even granting that the culpable aggressor's conduct is sufficient to constitute a culpable attack, McMahan questions how it can be the case that one is "justified" in killing the culpable aggressor as opposed to excused. That is, at best, the defender acts needlessly but understandably, and thus morality may deem him not blameworthy while still holding his action to be wrongful. After all, if the clerk or Harry knew the gun was unloaded, they could not act. Under such conditions, Greta and Ike would not be liable to defensive force. So how can not knowing a fact make something that was otherwise wrong, right?

I believe that this way of phrasing the problem begs a critical question about self-defense, and that is, whether self-defense has an epistemic component. It is certainly the case, as McMahan argues, that if you knew that the villain could not hurt you, you could not kill him. And one can see why McMahan views this as an excusing feature. If I take an umbrella believing it to be mine when it is in fact yours, then I do the wrong thing. However, because I believe the umbrella to be mine—I lack knowledge of a morally relevant fact—I do not appear to be blameworthy.⁷⁰

⁶⁷ Frowe, *supra* note 5, at 248 (defender does not wrong apparent culpable threat who is trying to kill him); Quong, *supra* note 12, at 519 (one forfeits one right not to be killed when one intentionally causes another person to believe he is trying to kill him).

⁶⁸ Ferzan, *supra* note 40, at 738 & n.69.

⁶⁹ McMahan, *supra* note 6, at 391.

⁷⁰ McMahan is an objectivist about justification. But even if one believes justification is about reasonable beliefs, there is still a question as to whether liability would turn on the defender's reasonable beliefs. After all, the question is whether the aggressor has forfeited a right, not whether the defender believes he has. These are different questions. For instance, a man who reasonably

There is, however, a critical difference between self-defense and the umbrella case. In the case of self-defense, the aggressor is the one who creates the very appearance of need. He creates the (mis)perception on the part of the defender. He is *responsible* for the defender's conduct.

This "morally relevant knowledge" problem arises because we believe that killing the aggressor is unnecessary. The defender has committed an unnecessary evil. Importantly, however, self-defense is not about committing the greater or lesser evil. It is about giving the defender a right to act on the basis of the threat that she perceives. *In no instance of self-defense will the defender be able to be 100% certain that she must act to protect herself.* She must act preemptively. She must respond to risk. And the perception of this risk has been culpably caused by the defender. Thus, he cannot stand to complain that she has acted on the very perception that he has created.

Importantly, this argument does not rely on the defender's perspective as its justificatory feature. Rather, the critical point is that the aggressor acted in a way that gave rise to the defender's belief. In my view, these words and/or actions are morally operative irrespective of whether the words stated are true. If Alice says to Betty, "you have my permission to borrow my car" and says this with a straight face, Alice has given Betty permission to drive the car, even if Alice says this as a joke. Alice can hardly claim that Betty violated her property right by using her car. Contract law is clear that one can incur a liability based on appearances. According to the Restatement (Second) of Contracts, section 19:

(2) The conduct of a party is not effective as a manifestation of his assent unless he intends to engage in the conduct and knows or has reason to know that the other party may infer from his conduct that he assents.⁷¹

Notice, then, that one may change one's rights and duties simply by behaving in a way that he has reason to know will allow the other party to infer that he is assenting to the contract. You do not have to intend to be bound to be bound. Thus the defender may render himself *liable* to be killed, even if he does not—from an omniscient perspective—*need* to be killed.⁷²

believes a woman is consenting cannot by his belief alone lead to her granting him a permission to have intercourse with her. Rather, the man is simply not culpable for his mistake.

⁷¹ RESTATEMENT (SECOND) OF CONTRACTS § 19 (1979).

⁷² Admittedly, I am unwilling to go so far as to say that if one negligently creates the appearance of a threat, one is also liable to be killed defensively. Although this reluctance requires a fuller defense than I wish to provide here, for now, I will simply say that I think that once we are not dealing with someone who is a threat, and we are just dealing with someone who is deficient in creating an appearance of a threat, the actor seems to me to have done too little to have forfeited his rights. And, I do think that we can require a higher standard for becoming liable to defensive force than for being liable to perform under a contract. My purpose at this point is merely to show that behaviors can alter rights and duties.

I suspect that McMahan would claim that I have missed what it means to be liable to defensive force. McMahan takes liability to self-defense to be that one can only be liable if harming him will serve a further purpose (preventing the harm to the defender).⁷³ These conditions do not obtain with bluffs because there is no actual harm to be prevented. McMahan may claim that I have conflated liability with forfeiture. I am skeptical, however, that we must embrace McMahan's conception of liability. As mentioned earlier, I have some reservations about what exactly self-defense theorists mean by liability. The Hohfeldian use of the term is that a liability is the correlative of a power.⁷⁴ In this sense, the attacker would have a *power* to alter the relationship between himself and the defender—he would have the ability to render himself liable to defensive force. This use of the term, however, would not, I take it, distinguish forfeitures, limited permissions, and McMahan's use of liability—all of them would be powers that the attacker could exercise under different conditions. Rather, the distinction between these would be drawn by how the rights and duties are altered—a boxer allows himself to be hit in a way quite different from the way that an attacker allows defensive force to be employed against him, and these two are quite different from a way that one changes one's duties to perform under a contract. For now, I will put this objection to the side. I take it that, at most, this amounts to the claim that I do not have a theory of liability, but a theory of limited forfeiture. I am prepared to live with that.

There is another potential objection to treating apparent threats as real ones. One might say that we can unravel this puzzle by separating provocation from aggression. Consider a case that does not give rise to a belief in the need to defend oneself. If during an argument, Carla yells at David, including racial epithets, insults to his manhood, and critiques of his child—particularly if she does so *in order for him to try to kill her*—and then he does so, she may not be not entitled to shoot him because she culpably created the situation.⁷⁵ Thus, she loses her defensive rights; she is not entitled to compensation; and she has effectively waived a right to complain about the attack.

One might therefore try to drive a wedge between apparent threats and actual ones, not by denying that the apparent threat alters rights and duties but by denying that the apparent threat alters *the same rights and duties*. The argument would run that the provocateur loses the right to defend and repel an attack, but does not

⁷³ MCMAHAN, *supra* note 7, at 8–9.

⁷⁴ WESLEY NEWCOMB HOHFELD, *FUNDAMENTAL LEGAL CONCEPTIONS AS APPLIED IN JUDICIAL REASONING* 36 (Walter Wheeler Cook ed., Greenwood Press 1978) (1919).

⁷⁵ This example is merely intended to create conceptual distance between provocateurs and aggressors. It is certainly questionable whether mere words should count for forfeiture of defensive rights. *See generally* State v. Riley, 976 P.2d 624, 628–29 (Wash. 1999) (discussing in dicta whether mere words should be enough). On the question of when earlier conduct affects the availability of a later defense generally, see Paul H. Robinson, *Causing the Conditions of One's Own Defense: A Study in the Limits of Theory in Criminal Law Doctrine*, 71 VA. L. REV. 1 (1985).

become liable to force or defensive force. That is, she is still wronged by the use of force, even if she lacks the right to defend herself against that force. She might also lose some compensatory rights, even though because she has not waived her right fully, the provoked attacker still wrongs her, and indeed, may still be criminally liable.

However, there is little basis for distinguishing the apparent threat from the actual one. We could note that from the defender's perspective, there is no difference; however, such an argument would surely be overinclusive, as from the defender's perspective there is no difference between an actual culpable threat and an apparent culpable threat, but as discussed above, it certainly matters that the aggressor is actually acting impermissibly in order for him to lose his right (though a defender might be fully excused so long as she believes the aggressor is culpable). But the view that I am stressing here is not fully aggressor-centered or defender-centered. Rather, the view of self-defense that I am urging sees liability as a communicative act that announces an attack and requires a defense. Not only does the defender hear the same thing when the attack is actual as when it is apparent, but also *the aggressor says the same thing*.

Indeed, reconsider *Involuntary Roulette*. Both Greta and Harry understand Greta's act in exactly the same way, and *neither* can predict whether Greta is an actual or apparent threat. How does it make sense for Greta's forfeiture of a right to turn on neither Harry's perception nor Greta's, but a fact that neither knows?

Ultimately, the sort of test we are after is one that looks to how much the attacker must do such that the conduct communicates to the defender, "I am going to hurt you unless you repel my attack."⁷⁶ It must be the sort of behavior that is intended to, or known to, cause fear (or, more accurately, a belief in the need to repel the attack). In other words, it is action which communicates the aggressor's intention to cause or risk causing the defender unjustified and unexcused harm. This action, whether true or not, is what the defendant must do culpably, and then, the aggressor is only liable to conduct designed to prevent that threat.

We are now in a position to reformulate our liability conditions:

A is liable to defensive force by *D* when

- (1) *A* forms an intention to purposefully, knowingly, or recklessly kill *D*, *A* lacks a justification or excuse, and *D* must kill *A* to prevent being killed himself, or
- (2) *A* purposefully, knowingly, or recklessly engages in conduct that he is aware may lead *D* to believe that (1) is true, and *A* lacks a justification or excuse for so doing.

⁷⁶ Elsewhere, I have argued that once the aggressor intends to cause the harm, any probability of success is sufficient to render him liable to defensive force. See Ferzan, *supra* note 40, at 744–45.

D. The Defensive Action Must Be Responsive to the Perceived Attack

This brings us to a final contour of liability. Jeff McMahan worries that if an attempt that cannot kill you is sufficient for liability, then there is no principled limitation on what sorts of culpable attempts render the defender liable.⁷⁷ Why not a culpable attempt ten years prior? He illustrates the problem with the following hypothetical case:

Aware that a villain plans to kill you, you begin to carry a gun. On one occasion you have the opportunity to empty the bullets from his gun and you do so. Immediately thereafter, he confronts you in an alley and tries to fire. As he continues to pull the trigger in frustration, you see that a second villain is preparing to shoot you from behind a narrow basement window (it is a tough neighborhood). Unable to flee in time and also unable to fire with accuracy through the tiny window, you can save yourself only by shooting the first villain, causing him to slump in front of the window, thereby blocking the second villain's line of fire.⁷⁸

In this case, the first villain is trying to kill you, and you now need him to die to save yourself. May you kill him? If the answer is yes, then notice how we have severed liability from need. Once one has become liable to defensive force, one becomes liable to any sort of defensive force as needed. And, then, McMahan's argument goes, because there is no real threat, there is no principled limitation to when you cease to be liable to any sort of needed defensive response.⁷⁹

I am unwilling to follow McMahan down this primrose path. My answers to the question of whether you may kill the first villain are no and maybe. No, as a matter of self-defense. Maybe, as a matter of agent-relative permissions. To just gesture at the latter analysis, I believe that the means principle constrains the sorts of actions that we may consequentially justify. Thus, even if killing a culpable person is a lesser evil than dying oneself, one could not use a culpable person. On the other hand, I take it that we may be excused (or there may be an agent relative permission) that entitles us to act as a person of reasonable firmness would.⁸⁰ It may be too much for morality to ask of me not to use the first villain, and particularly, not to do so given that he is trying to kill me.

Let us return to liability. Is the first villain *liable* to be killed by me? Here, I think the answer is no. Recall that liability is not complete forfeiture. Once someone is liable to defensive force, you don't get to carve up his organs and the like. This is liability to defensive force. It is: "Have you done something that

⁷⁷ McMahan, *supra* note 6, at 392–93.

⁷⁸ *Id.* at 391.

⁷⁹ *Id.* at 392–93.

⁸⁰ See ALEXANDER & FERZAN, *supra* note 48, at 130 (discussing when culpable persons may be used); Ferzan, *supra* note 17, at 510.

allows me to repel your attack?" It is not: "Have you acted in such a way that you are a bad person so I can use you willy-nilly?"

The conceptual and normative structures of self-defense are interrelated. Self-defense is about defense. It is not about aggression or things we are permitted to do to others. Self-defense repels aggression. Liability conditions are those behaviors by an aggressor that give rise to the defender's permission to repel the attack, even when repelling the attack requires harming the aggressor. In the same way that my permission to drive my car is not permission for you to sell its parts, an aggressor's attack is not a wholesale moral forfeiture but a limited permission to harm another only for the purpose of repelling that attack.⁸¹

Critics of forfeiture views have consistently pointed to forfeiture's seemingly limited explanatory force.⁸² How can an aggressor forfeit a right to life but regain it as soon as she desists in her attack?⁸³ How can this right be forfeited in limited ways so as only to permit repelling the attack but not harvesting the aggressor's organs?⁸⁴ However, there is nothing particularly odd about seeing our powers to change our rights as being fairly nuanced. A defendant who kills a star witness to prevent her from testifying against him may forfeit his right to confrontation but he does not forfeit his right to a fair trial.⁸⁵ With self-defense, by attacking or knowingly appearing to attack the defender, the aggressor gives the defender a right to repel the attack.

Still, one might find it odd to think that there is an "uptake" element to self-defense. Why should the aggressor's liability turn on whether the defender is responding to that liability? Should not the responsiveness be a question of permissibility? Consider:

Unknown Attack. Ned, intending to kill Nedda, points a plastic gun at her. Although the gun appears to be a toy gun, it is real. Nedda believes the gun to be a toy, but hoping to convince a jury that she believed the gun to be real, uses this as an opportunity to kill Ned as she had already planned to do that day. Nedda kills Ned.

⁸¹ McMahan himself makes a similar claim, "The right against attack is instead forfeited only in relation to certain persons acting for certain reasons in a particular context." McMAHAN, *supra* note 7, at 10.

⁸² Judith Jarvis Thomson, *Self-Defense and Rights*, in RIGHTS, RESTITUTION, & RISK: ESSAYS IN MORAL THEORY 33 (1986).

⁸³ *Id.* at 34.

⁸⁴ *Id.*

⁸⁵ *Giles v. California*, 554 U.S. 353 (2008) (holding defendant only forfeits Confrontation right when his wrongdoing that results in the witness' unavailability was intended to prevent a witness from testifying).

Criminal law theorists call this a case of "unknowing justification."⁸⁶ The question for us is whether Ned is liable to be killed defensively, and therefore, whether Nedda wrongs Ned by killing him. On one side of the argument, we might want to say that Ned is liable to be killed. He is culpably trying to kill her, and Nedda's killing him is the only thing that actually stops him. On the other hand, we can point to the fact that Nedda is not responding defensively. She is aggressing. The question then is whether (1) Ned is liable to be killed defensively but that is not what happened so Nedda wronged Ned, or (2) Ned is liable to be killed defensively and that is what happened but because Nedda did not know this she lacked a belief that her conduct was necessary and therefore acted culpably but not objectively wrongfully. That is, once an aggressor is liable to defensive force, must that force be intended or known to be defensive? Or it sufficient that the force stops the attack?

The defender must be acting to repel the attack. Morally, there is simply no reason for us to prefer Ned over Nedda simply because Ned's attack occurred before Nedda's. Indeed, it gets us into questions such as whether Nedda started her plan earlier, such that we can figure out who was "first." Those can't be the correct questions. The correct questions have to be about why the attacker has forfeited his moral claim. I take it that even would-be killers do not forfeit their claims against being killed, used, and the like. They forfeit their claims about actions that are responsive to their attacks.⁸⁷ That is, attackers forfeit their moral claims for defenders to act for defensive reasons, not simply to harm them in a way that fortunately and unintentionally stops their attacks.

For this reason, we see that there is a principled answer to the limits of one's liability for a culpable action. When one becomes liable to defensive force, one is only liable to force that is intended to repel one's attack. One is not liable to other force, even if that force will assist the defender.

Let us now consider another complication. When the defender acts, what must she believe about the attacker's culpability? Consider:

Known Threat/Unknown Culpability. Shmed intends to kill Shmedda by pressing a button on his cell phone that will detonate a bomb in close proximity to Shmedda who lies in Shmed's line of sight but beyond shouting distance. Shmedda sees that Shmed is about to press the button,

⁸⁶ See Ferzan, *supra* note 9, at 244 ("Unknowing justifications are those instances in which the actor lacks the so-called justifying beliefs, but the facts support a justification."). The famous debate over the treatment of these cases began with an exchange between George Fletcher and Paul Robinson. See Paul Robinson, *A Theory of Justification: Societal Harm as a Prerequisite to Criminal Liability*, 23 UCLA L. REV. 266 (1975); George P. Fletcher, *The Right Deed for the Wrong Reason: A Reply to Mr. Robinson*, 23 UCLA L. REV. 293 (1975).

⁸⁷ I believe it is sufficient that the defender know that he is responding defensively. Mixed motive cases qualify as defensive.

believes (correctly) that the result will be her death, but also believes (mistakenly) that Shmed is unaware that this action imperils her.⁸⁸

Here, Shmedda is acting defensively and she is acting against the threat that she perceives. However, she does not believe that Shmed is culpable. Should this matter? After all, if Shmedda were correct and Shmed was an innocent aggressor, then Shmedda would only be able to offer an excuse or a personal justification for defending herself. Still, with respect to whether Shmed has forfeited his right, it seems that he has because he is presenting a threat, and Shmedda is responding to it. Hence, as a matter of Shmedda's criminal liability and blameworthiness, this will turn on her beliefs with respect to Shmed. But as to whether Shmed has forfeited his rights, and thus as to whether he is able to fight back or enlist aid, he has forfeited his rights, and he may not fight back or enlist aid, because he is in fact culpable and Shmedda is responding defensively to his perceived (and in this case, actual) threat.

With this feature in place, we can now see the structure of self-defense. When an aggressor, with a culpable mind, acts in a way that gives rise to a belief that he will kill the defender, he forfeits his moral claim against the defender responding to the perceived threat by repelling the attack.

III. CONCLUSION

Although our use of the term, "self-defense," covers myriad actions that respond to apparent threats, one important category of self-defensive action is liability-based defense. When an aggressor is liable to defensive force, he is not wronged by force that is designed to repel his attack.

Liability to Defensive Force:

If:

(1a) The aggressor forms an intention to purposefully, knowingly, or recklessly kill the defender, the aggressor lacks a justification or excuse, and the defender must kill the aggressor to prevent being killed himself, or

(1b) The aggressor purposefully, knowingly, or recklessly engages in conduct that he is aware may lead the defender to believe that (1a) is true, and the aggressor lacks a justification or excuse for so doing;

Then:

(2) The aggressor is not wronged by force employed by the defender that is intended to repel the perceived attack.

⁸⁸ I thank Mitch Berman for this example and for raising this issue.

